ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.17, No.8, August 2006, pp.1775−1784
DOI: 10.1360/jos171775

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

流测量中基于测量缓冲区的时间分层分组抽样*

王洪波+, 韦安明, 林 宇, 程时端

(网络与交换技术国家重点实验室(北京邮电大学),北京 100876)

Time Stratified Packet Sampling Based on Measurement Buffer for Flow Measurement

WANG Hong-Bo+, WEI An-Ming, LIN Yu, CHENG Shi-Duan

(State Key Laboratory of Networking and Switching Technology (Beijing University of Posts & Telecommunications), Beijing 100876, China)

+ Corresponding author: Phn: +86-10-62264949 ext 201, Fax: +86-10-62283412, E-mail: hbwang@bupt.edu.cn

Wang HB, Wei AM, Lin Y, Cheng SD. Time stratified packet sampling based on measurement buffer for flow measurement. *Journal of Software*, 2006,17(8):1775−1784. http://www.jos.org.cn/1000-9825/17/1775.htm

Abstract: Although NetFlow is widely deployed for traffic measurement, the sampling method of Netflow has shortcomings: it consumes excessive router resource during flooding attacks; selecting a suitable static sampling rate is difficult because no single rate gives the right tradeoff between resource consumption and accuracy for all traffic mixes. An easily-implemented packet sampling method is presented in this paper, which samples a fixed number of packets in the constant period with measurement buffer. The method can automatically adapt the sampling rate to traffic variety and provide the controllability of resource consumption. Theoretical analyses demonstrate that the new method can provide unbiased estimation with certain relative standard deviation bound. Experiments are also conducted with the real network traces. Results show that the proposed method can achieve simplicity, adaptability and controllability of resource consumption without sacrificing accuracy compared with other sampling methods.

Key words: traffic measurement; network monitoring; IP flow; packet sampling; NetFlow

摘 要: NetFlow是流测量中广泛应用的解决方案,但NetFlow的抽样方法存在一定的缺陷:泛洪攻击时消耗路由器过多的资源;用户很难选择适合所有流量组成情况的静态抽样率,以平衡资源消耗量和准确率.提出了一种易于实现的分组抽样方法.该方法利用测量缓冲区对定长时间内到达的分组进行固定数量的抽样,既可以使抽样率自适应于流量变化,又可以控制资源的消耗.证明了抽样估计的无偏性,并推导出估计值相对标准差的理论上界.实验结果表明,与已有方法相比,该方法在具有简单性、自适应性及资源可控性的同时不会失去准确性.

关键词: 流量测量;网络监控;IP 流;分组抽样;NetFlow
中图法分类号: TP393 文献标识码: A

* Supported by the National Natural Science Foundation of China under Grant Nos.90604019, 60472067, 60502037 (国家自然科学基金); the National Grand Fundamental Research 973 Program of China under Grant Nos.2003CB314806, 2006CB701306 (国家重点基础研究发展规划(973)); the China Next Generation Internet Project under Grant No.CNGI-04-8-1D (国家 CNGI 项目)

Received 2005-08-23; Accepted 2006-01-09

流量测量(traffic measurement)是网络监测、管理和控制的基础.目前有 3 种方法用于流量测量:一种是使用 SNMP(simple network management protocol)统计数据来获取流量信息,但是它只能提供粗粒度的流量信息,不能满足深入分析的要求;另一种是通过采集流经链路的分组来进行流量测量,它能够支持广泛的分析和应用,但可扩展性较差;再一种是流*(flow[1])级别的测量,它既能提供详细的流量信息,又具备一定的可扩展性,因而受到广泛的关注并得到大量应用.为此,IETF(Internet Engineering Task Force)组织在流测量(flow measurement)方面做了大量的工作[2,3].主流商用路由器大都支持流测量功能,如 Cisco 的 NetFlow[4]就是广泛使用的流测量工具.目前,NetFlow 已经成为流测量的主要工具,诸如主要业务、主要用户、流量矩阵等信息都可由此获得.

然而,链路带宽的快速增长和网络流量的急剧膨胀使得流测量在高速网络中仍然面临着可扩展性挑战.为此,人们提出基于抽样的流测量:首先对分组进行抽样,然后对分组样本进行基于流的统计,从而保持流测量的可扩展性.Cisco 在 NetFlow 中引入抽样机制形成随机抽样 NetFlow(random sampled NetFlow,简称 RSNF)[5],以适应网络的高速化.虽然如此,RSNF 的抽样方法仍然存在一些缺点[6]:(1) 抽样率**(sampling rate)需人工配置,且在测量时不可变,造成使用不便.高抽样率会因过多消耗资源而引起路由器性能的降低;低抽样率因获取样本数量过少、总体信息缺失过多而造成分析结果误差加大.如何选择合适的抽样率往往使用户处于两难境地.(2) 所用资源随流量波动而变化,缺乏资源保护功能.NetFlow 主要消耗路由器的处理器资源、流记录所占用的存储器(“流 cache”)资源和输出流记录时的带宽资源.由于资源消耗量与分组或流记录数成正比,因此在高负载条件下,尤其是当泛洪攻击(flooding attack)等事件所引起的网络流量突然上升时,资源会被过度占用,从而影响路由器的正常转发功能.

这就需要研究新的抽样方法,以满足 3 个目标:(1) 简单性:用户操作的简单性,减轻用户设置参数时所面临的繁复工作,减少人工操作;实现的简单性,抽样方法应易于理解、实现简单;(2) 自适应性:抽样方法应能根据流量的变化动态调整抽样率;(3) 资源消耗的可控性:NetFlow 对资源的占用不能影响其他功能的正常使用.

在网络测量领域,围绕分组抽样方法已有许多研究工作[7−10].但这些工作都是为了解决不同的应用问题而做的,并不能应用于流测量.近年来,与流测量相关的抽样方法引起了相当的重视.文献[11]以保证大流的估计准确性为目标,使用自回归模型预测分组总数来动态调整分组抽样率,但并未讨论资源使用的保护.Duffield 等人在文献[12]中提出了一种非均匀流抽样方法,其目的是在流测量设备输出流记录时,通过对流记录的抽样来适应资源的限制.而本文及其他文献则是在测量点对分组进行抽样.文献[6]首次指出了 NetFlow 中存在的问题,并提出了一种具有自适应调整抽样率的 Adaptive NetFlow(ANF).其核心思想是:测量开始时先以 CPU 所能接受的最大抽样率进行试探性抽样,在测量过程中不断根据“流 cache”资源的使用量及其中流记录的分组数分布情况来降低抽样率,并以新的抽样率来重正化(renormalization)已有流记录,而丢弃重正化后分组数为 0 的流记录,以便为新的流记录腾出空间.该方法的优点是能够保证每个测量时间段内输出的流个数最大值是确定的.其缺点是:(1) 最大抽样率是根据链路在最极端(即所有分组为最小分组(40Byte),链路为最大利用率)的条件下计算的,而通常情况下,链路使用率远小于此,因而 CPU 资源不能得到充分的利用;(2) 在测量时间段内抽样率只能递减,即使“流 cache”有足够的空间,抽样率也无法再提高,这将影响估计准确率;(3) 重正化操作需要与分组处理并行进行,而且必须保证重正化释放流记录的速率要大于新流记录的创建速率,这都增加了方法实现的复杂度.

本文针对 NetFlow 抽样方法中的现存问题提出了一种基于测量缓冲区的时间分层分组抽样方法.它通过引入测量缓冲区对定长时间内到达的分组进行固定数量的抽样,既达到了抽样率自适应于流量变化的目的,又可以控制对测量设备资源的消耗,而且该方法易于实现.

本文第 1 节描述我们提出的抽样及估值方法.第 2 节对估值进行理论分析.第 3 节使用实际互联网数据进行实验验证和方法比较.最后一节总结全文.

---

* 流是一系列具有共同属性的分组集合,且这些分组的前后到达时间间隔小于某个给定值.

** 抽样率为样本数与总体的比值.

1 基于测量缓冲区的时间分层分组抽样

1.1 基本思想

针对引言中的 3 个改进目标,我们的改进的基本思想如下:

设立测量缓冲区以分离分组转发与测量功能:转发 IP 分组是路由器的主要功能,分组的线速转发是路由器设计的主要技术指标之一,而线速转发就要求路由器在最小的分组(40Byte)到达时间间隔完成每一个分组的处理.NetFlow 功能中对分组的处理是与 IP 分组的转发处理结合在一起的,也就是说,当路由器以线速转发 IP 分组的同时,线速完成 NetFlow 的分组处理.实际上,测量任务并没有像转发分组这样高的实时要求.另外,随着需求的不断增加,测量、统计等功能也不断增多,分组的非转发处理时间需求将随之加大,从而增大了线速转发的设计难度.将测量功能与转发功能分开,可以减小新增功能对已有关键功能的性能影响,这也便于测量功能的扩展.因此,我们提出在系统中设计一个测量缓冲区,当分组到达时,若它被作为样本抽取,则将该分组头部的相关信息复制到缓冲区中,由独立运行的测量进程完成统计功能.

在定长时间内抽样固定数量的分组,以提供资源可控性及抽样率的自适应性:网络的流量是随时间波动变化的,为了控制流测量所使用的资源,根据可提供的最大资源情况在一定的时间内随机抽取固定数量的分组.这样既可以使资源占用有一个上界,又可以使资源的使用趋于平稳,从而达到资源可控的目的.另一方面,由于在定长时间段内抽取固定数量的分组,当分组到达速率高时,抽样率降低;反之,抽样率升高,从而达到抽样率对流量变化的自适应.这种抽样率的调整是根据当前流量情况作出的,具有实时性,而不是像预测反馈方法那样根据当前状态预测之后的抽样率,克服了其他方法中存在的预测错误及抽样率调整延迟等缺点.另外,由于不需要抽样率预测以及与抽样率调整相关的操作(如文献[6]中的重正化)而使得方法易于实现及使用.

1.2 抽样设计

NetFlow 使用 4 个规则来结束一个流并输出该流记录:(1) 该流新到达的分组为 TCP 数据,且 TCP 首部的 FIN 或 RST 标志位被置位;(2) 在一定的时间周期内(默认为 15 秒,用户可配置),没有属于该流的分组到达;(3) 流记录存活时间超时(默认为 30 分钟,用户可配置);(4) 当“流 cache”无空间时.

实际应用中往往需要计算不同时间段内的流量,而当根据 NetFlow 的流记录中的起止时间来计算各个时间段内的流量时会引入误差[6].因此,文献[6]提出把整个 NetFlow 运行时间分成较短的等长测量时间段,在测量时间段内不结束任何流,而在测量时间段结束时结束“流 cache”中的所有流.出于同样的考虑,本文也采用划分测量时间段的方法,与之不同的是,在一个测量时间段内,我们仍保留 NetFlow 关于流的结束规则以提高“流 cache”的利用率***,并保证总有空间提供给新的流记录,在每个测量时间段结束时结束所有流.测量时间段的长度可以作为可配置参数,在本文的实验中,与文献[6]相同,采用 1 分钟.由于每个测量时间段内的抽样过程都是一样的,因此后文讨论的抽样都是在一个测量时间段范围内进行.

基于上一节中的设计思想,一个测量时间段内的抽样过程为:划分一个测量时间段为 m 个等长度的子测量时间段;对每个子测量时间段,把在此期间内到达的所有分组视为总体并使用测量缓冲区进行简单随机抽样,抽取 n 个分组进行统计处理,在统计处理时,使用此子测量时间段的抽样率进行估计.划分子测量时间段实际上是对整个测量时间段内到达的分组按到达时间范围进行了分簇,每个簇内进行单独的抽样,这实质上是一种时间分层抽样,又由于在每个分层中的抽样是基于测量缓冲区进行的,因而我们称这种抽样方法为基于测量缓冲区的时间分层分组抽样(time stratified packet sampling based on measurement buffer,简称 TSPSBMB).

1.2.1 子测量时间段内的简单随机抽样

由于测量缓冲区空间大小有限,不可能把子测量时间段内到达的所有分组都保存下来再进行抽样,必须在分组到达的过程中进行抽样,而在子测量时间段结束前到达的分组总数是未知的.因此,此时的问题是如何在事先未知总数的分组中抽样固定个数的样本放到测量缓冲区中.此问题可用文献[14]中的“蓄水池抽样算法”

*** 事实上,流持续时间服从重尾分布,大部分流具有较短的持续时间和较短的最大分组间隔时间[13].

(random sampling with a reservoir)来解决.其基本思想是:先把前 n 个到达的分组放到测量缓冲区中作为候选的抽样样本,当第 $t(t>n)$个分组到达时,以 n/t 概率成为候选样本并随机替换缓冲区中的一个候选样本,这样做一直到子测量时间段结束.最后,测量缓冲区中的分组就构成了 n 个简单随机抽样样本.由于要对每个分组进行随机数的计算,其时间复杂度为 $O(N)$,其中 N 为在子测量时间段内到达的分组总数.文献[14]中的改进算法 Z 的时间复杂度可降为 $O(n(1+\log(N/n)))$.

1.2.2 测量缓冲区设计

对于流量测量任务来说,在测量缓冲区中只需存放分组头部的某些域及其他信息.例如:分组头部中的域有服务类型(type of service,简称 TOS)、总长度(total length)、协议(protocol)、源站 IP 地址、目的站 IP 地址、源端口号、目的端口号、TCP 的标志比特(flag bits);其他信息有分组到达时的时间戳,此信息可用来计算流的起止时间等.

由于 n 个分组抽样样本只有在子测量时间段结束时才最终形成,因而只有在此时才能开始对缓冲区中的分组样本进行处理.与此同时,下一个子测量时间段的分组将陆续到达,因此,需要有更多的缓冲区来保存分组.这可以通过循环队列来解决.为简单起见,我们给出较为保守的缓冲区设计:

整个缓冲区由 A 与 B 两部分组成,每部分可存放 n 个分组信息;

对子测量时间段按时间顺序进行编号,i=1,2,3,…,当 i 为奇数时,从链路到达且需要缓存的分组,相应信息进入缓冲区 A,而测量进程从缓冲区 B 读取抽样样本(当 i=1 时,B 中无抽样样本);当 i 为偶数时,从线路到达且需要缓存的分组,相应信息进入缓冲区 B,而流测量模块从缓冲区 A 读取抽样样本.如图 1 所示.

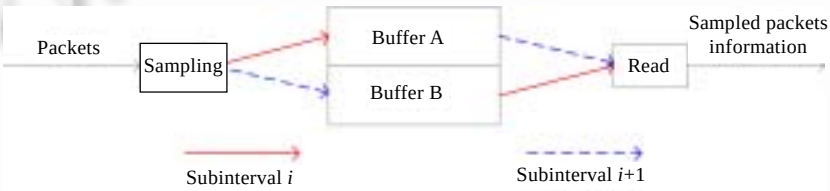

Fig.1 Measurement buffer structure

图 1 测量缓冲区结构

对于每个子测量时间段,由于要缓存所有前 n 个到达的分组,即必须以线速存储前 n 个分组,所以缓冲区需要用静态随机访问存储器 SRAM(static random access memory)来实现.对于能容纳 12 000 个分组信息的缓冲区来说,如果每个分组信息需要 21Bytes,则最多需要 3.85Mbit 的 SRAM.根据当前的半导体技术,提供这种大小的 SRAM 是可行的.

1.3 估值方法

当子测量时间段结束时,在此子测量时间段内到达的分组总数是可确定的(通过计数器计数),而抽取的样本数是定值 n,因而可以求出此子测量时间段的实际抽样率.在对一个子测量时间段的 n 个样本进行统计时,将属于某个流的所有样本分组的字节总数(分组总数)除以此子测量时间段的实际抽样率作为此子测量时间段内该流的字节数(分组数)的估计值,然后加到流记录的相应统计字段中.当整个测量时间段结束时,流记录中相应统计字段就是其估计值.

形式化描述如下:假设测量时间段被均分成 m 个子测量时间段,分别记为 $t_1,t_2,t_3,\ldots,t_m$;把测量时间段内出现的每一个流记为 $f_k(k=1,2,\ldots)$.在子测量时间段 $t_i(i=1,2,\ldots,m)$期间到达的分组数记为 N_i,当时间段结束时,由于样本数是 n ,因此该子测量时间段的抽样率为 n/N_i.在子测量时间段 t_i 内属于流 f_k 的分组数记为 N_i^k ,其中第 j 个分组记为 $p_{ij}^k(j=1,2,...,N_i^k)$,此分组的字节数记为 x_{ij}^k ;抽样后属于流 f_k 的样本分组数记为 n_i^k ;其中第 j 个分组记为 $q_{ij}^k(j=1,2,...,n_i^k)$,此分组的字节数记为 X_{ij}^k .

在子测量时间段 t_i 内,流 f_k 的实际字节数 $x_i^k=\sum_{j=1}^{N_i^k}X_{ij}^k$ 的估计值为

$$\hat{x}_i^k=\frac{N_i}{n}\sum_{j=1}^{n_i^k}X_{ij}^k \tag{1}$$

在子测量时间段 t_i 内,流 f_k 分组数 N_i^k 的估计值为

$$\hat{N}_i^k=\frac{N_i}{n}n_i^k \tag{2}$$

在整个测量时间段内,流 f_k 的实际字节数 $x^k=\sum_{i=1}^{m}x_i^k$ 的估计值为

$$\hat{x}^k=\sum_{i=1}^{m}\hat{x}_i^k \tag{3}$$

在整个测量时间段内,流 f_k 的分组数 $N^k=\sum_{i=1}^{m}N_i^k$ 的估计值为

$$\hat{N}^k=\sum_{i=1}^{m}\hat{N}_i^k \tag{4}$$

2 理论分析

在实际流量分析中,经常关心流量各组成部分的总量及其比例情况(例如求来自某一个网络的 HTTP 应用的总字节数),这时,需要对多个流的估计值进行求和.如果单个流的估计值具有无偏性,则多个单个流的估计误差在求和过程中不会有累积效应,因此在抽样设计中,估计值的无偏性是重要的.本节证明 $\hat{x}^k$ 和 $\hat{N}^k$ 的无偏性,并推导它们的相对标准差的上界.

先给出几个记号:对于子测量时间段 t_i 内的流 f_k,记 $\mu_i=\frac{1}{N_i^k}\sum_{j=1}^{N_i^k}x_{ij}^k$, $\sigma_i^{\,2}=\frac{1}{N_i^k}\sum_{j=1}^{N_i^k}(x_{ij}^k-\mu_i)^2$;在整个测量时间段内,记流 f_k 的分组平均字节数为 $\mu^k=\frac{x^k}{N^k}$,N 为测量时间段内所有流的分组总数.

引理 **1**. 在子测量时间段 t_i 内,具有 N_i^k 个分组的流 f_k 被抽样的分组数 n_i^k 是服从两项分布 $B(N_i^k,n/N_i)$ 的随机变量,n_i^k 的期望及方差分别为 $E(n_i^k)=N_i^k\frac{n}{N_i}$, $Var(n_i^k)=\frac{n}{N_i}\left(1-\frac{n}{N_i}\right)N_i^k$.

证明:由于在子测量时间段 t_i 内进行的是简单随机抽样,总共有 N_i 个分组,每个分组被抽样的概率为 n/N_i,因而属于流 f_k 的 N_i^k 个分组中被抽样的分组数 n_i^k 是服从两项分布的随机变量,即 $n_i^k\sim B(N_i^k,n/N_i)$,由两项分布性质可知, $E(n_i^k)=N_i^k\frac{n}{N_i}$, $Var(n_i^k)=\frac{n}{N_i}\left(1-\frac{n}{N_i}\right)N_i^k$. □

定理 **1**. 任何一个流 f_k,它的分组数的估计值 $\hat{N}^k$ 具有无偏性.

证明:由式(2)及引理 1 可得, $E(\hat{N}_i^k)=E\left(\frac{N_i}{n}n_i^k\right)=\frac{N_i}{n}E(n_i^k)=N_i^k$.

又由式(4)有: $E(\hat{N}^k)=E\left(\sum_{i=1}^{m}\hat{N}_i^{\,k}\right)=\sum_{i=1}^{m}E(\hat{N}_i^{\,k})=\sum_{i=1}^{m}\hat{N}_i^{\,k}=N^k$,

所以, $\hat{N}^k$ 具有无偏性. □

定理 **2**. 任何一个流 f_k,它的字节数的估计值 $\hat{x}^k$ 具有无偏性.

证明:在子测量时间段 t_i 内,流 f_k 的 n_i^k 个分组被抽样,每个抽样分组对应字节数 $X_{ij}^k(j=1,2,...,n_i^k)$ 是随机变量,由简单随机抽样的性质[15]可知: $E(X_{ij}^k)=\mu_i$;再由式(1)有:

$$E(\hat{x}_i^k \mid n_i^k) = E\left(\frac{N_i}{n}\sum_{j=1}^{n_i^k} X_{ij}^k\right) = \frac{N_i}{n} n_i^k \mu_i \tag{5}$$

从而, $E(\hat{x}_i^k) = E(E(\hat{x}_i^k \mid n_i^k)) = E\left(\frac{N_i}{n} n_i^k \mu_i\right) = \frac{N_i}{n} E(n_i^k)\mu_i = N_i^k \mu_i = x_i^k$ (倒数第 2 个等式由引理 1 可得),

再由式(3)得: $E(\hat{x}^k) = E\left(\sum_{i=1}^{m} \hat{x}_i^{\,k}\right) = \sum_{i=1}^{m} E(\hat{x}_i^{\,k}) = \sum_{i=1}^{m} x_i^k = x^k$,

所以, $\hat{x}^k$ 具有无偏性. □

定理 **3**. 流 f_k 的分组数估计值 $\hat{N}^k$ 的相对标准差具有上界 $1\Big/\sqrt{n\frac{N^k}{N}}$.

证明:由式(2)及引理 1,流 f_k 在子测量时间段 t_i 内的分组数估计值的方差为

$$Var(\hat{N}_i^k) = Var\left(\frac{N_i}{n} n_i^k\right) = \frac{N_i^2}{n^2} Var(n_i^k) = \frac{N_i^2}{n^2}\left(\frac{n}{N_i}\left(1-\frac{n}{N_i}\right)N_i^k\right) = \frac{N_i}{n}\left(1-\frac{n}{N_i}\right)N_i^k < \frac{N_i}{n} N_i^k .$$

由于子测量时间段之间的抽样相互独立,所以 $\hat{N}_i^{\,k}(i=1,2,...,m)$ 是相互独立的随机变量,再由式(4),流 f_k 在整个测量时间段内分组数估计值的方差 $Var(\hat{N}^k) = Var\left(\sum_{i=1}^{m} \hat{N}_i^{\,k}\right) = \sum_{i=1}^{m} Var(\hat{N}_i^{\,k}) < \sum_{i=1}^{m} \frac{N_i}{n} N_i^k = \frac{1}{n}\sum_{i=1}^{m} N_i N_i^k$; 注意到 $N_i>0$ 且 $N_i^k \ge 0$,所以有: $\sum_{i=1}^{m} N_i N_i^k \le \sum_{i=1}^{m} N_i \sum_{i=1}^{m} N_i^k = NN^k$;由上所述可知: $Var(\hat{N}^k) < \frac{NN^k}{n}$.因而,流 f_k 分组数估计值的相对标准差 $\frac{\sqrt{Var(\hat{N}^k)}}{N^k} < 1\Big/\sqrt{n\frac{N^k}{N}}$. □

定理 3 说明:对一个流的分组数估计值来说,其相对标准差的上界由测量缓冲区大小(n)和该流分组数在总流量中所占的比例(N^k/N)所决定,而与抽样率无关;一个流的分组数所占比例越大,它的估计误差的最大值就越小,这是符合直觉的;对所有流而言,测量缓冲区空间越大,标准差越小,这意味着可以通过测量缓冲区大小来控制误差的上界.值得注意的是:估计误差仍与抽样率有关——抽样率越小,误差越大,会更接近于这个上界.这由证明过程可知.

定理 **4**. 流 f_k 字节数估计值 $\hat{x}^k$ 的相对标准差具有上界 $1\Big/\sqrt{n\frac{N^k}{N}\frac{\mu^k}{b_{\max}}}$,其中 $b_{\max}$ 为 IP 网中分组最大字节数.

证明:因在子测量时间段 t_i 内进行的是简单随机抽样,所以,流 f_k 在子测量时间段 t_i 内字节数估计值的方差为

$$\begin{aligned} Var(\hat{x}_i^k \mid n_i^k) &= Var\left(\frac{N_i}{n}\sum_{j=1}^{n_i^k} X_{ij}^k\right) \\ &= Var\left(\frac{N_i}{n} n_i^k \frac{1}{n_i^k}\sum_{j=1}^{n_i^k} X_{ij}^k\right) \\ &= \left(\frac{N_i}{n} n_i^k\right)^2 Var\left(\frac{1}{n_i^k}\sum_{j=1}^{n_i^k} X_{ij}^k\right) \\ &= \left(\frac{N_i}{n} n_i^k\right)^2 \frac{\sigma_i^2}{n_i^k}\left(1-\frac{n_i^k-1}{N_i^k-1}\right) \\ &< \frac{N_i^2}{n^2} n_i^k \sigma_i^2 . \end{aligned}$$

上面的第 4 个等式利用简单随机抽样样本均值的方差公式[15]得到.再由式(5)及引理 1 有:

$$\begin{aligned}Var(\hat{x}_i^k) &= E(Var(\hat{x}_i^k \mid n_i^k)) + Var(E(\hat{x}_i^k \mid n_i^k)) \\ &< E\left(\frac{N_i^2}{n^2} n_i^k \sigma_i^2\right) + Var\left(\frac{N_i}{n} n_i^k \mu_i\right) \\ &= \frac{N_i^2}{n^2} E(n_i^k)\sigma_i^2 + \frac{N_i^2}{n^2} Var(n_i^k)\mu_i^2 \\ &= \frac{N_i}{n} N_i^k \sigma_i^2 + \frac{N_i}{n} N_i^k \left(1 - \frac{n}{N_i}\right)\mu_i^2 \\ &< \frac{N_i}{n} N_i^k (\sigma_i^2 + \mu_i^2).x.\end{aligned}$$

而 $\sigma_i^2 + \mu_i^2 = \frac{1}{N_i^k}\sum_{j=1}^{N_i^k}(x_{ij}^k)^2$,因而 $Var(\hat{x}_i^k) < \frac{N_i}{n}\sum_{j=1}^{N_i^k}(x_{ij}^k)^2$.

又因为对所有的 i,j,k 都有 $x_{ij}^k \le b_{\max}$,所以 $Var(\hat{x}_i^k) < \frac{b_{\max} N_i}{n}\sum_{j=1}^{N_i^k} x_{ij}^k$.

由于子测量时间段之间的抽样相互独立,所以 $\hat{x}_i^k (i = 1,2,...,m)$ 是相互独立的随机变量,因而流 f_k 在整个测量时间段内字节数估计值的方差: $Var(\hat{x}^k) = Var\left(\sum_{i=1}^{m}\hat{x}^k\right) = \sum_{i=1}^{m} Var(\hat{x}_i^k) < \frac{b_{\max}}{n}\sum_{i=1}^{m}\left(N_i \sum_{j=1}^{N_i^k} x_{ij}^k\right)$.

又因为 $\sum_{i=1}^{m}\left(N_i \sum_{j=1}^{N_i^k} x_{ij}^k\right) \le \left(\sum_{i=1}^{m} N_i\right)\left(\sum_{i=1}^{m}\sum_{j=1}^{N_i^k} x_{ij}^k\right) = Nx^k$, 所以 $Var(\hat{x}^k) < \frac{b_{\max} N x^k}{n}$.

因而,流 f_k 字节数估计值的相对标准差为 $\frac{\sqrt{Var(\hat{x}^k)}}{x^k} < \sqrt{\frac{b_{\max} N}{n x^k}} = \sqrt{\frac{b_{\max} N}{n \mu^k N^k}} = 1\Big/\sqrt{n\frac{N^k}{N}\frac{\mu^k}{b_{\max}}}$. □

定理 4 与定理 3 的结论是类似的.对于一个流的字节数估计值而言,其相对标准差的上界与抽样率无关,影响因素为测量缓冲区大小、该流分组数在总流量中所占的比例以及流 f_k 的分组平均字节数与 IP 网分组最大字节数的比值($\mu^k/b_{\max}$).同样,也可以通过测量缓冲区大小来控制字节数估计误差的上界.

3 实验验证

本节利用互联网实际数据对抽样方法进行实验验证.数据由 CAIDA(Cooperative Association for Internet Data Analysis)[16]组织在美国某 ISP(Internet service provider)的 OC-48 骨干链路上采集所得,本文使用分布于 1 小时内不同时间的 8 段 1 分钟数据,数据描述见表 1.

Table 1 Summary of experiment data

表 1 实验数据概要

Data source	Date	Start time (UTC-8)	Bits/s (M)	Pkts/s (k)	Flow/s
OC48 link A (CAIDA)	January 15, 2003	10:04	298.9	55.1	6 741
		10:09	311.0	55.8	6 655
		10:14	320.1	56.4	6 586
		10:19	306.5	55.3	6 659
		10:24	324.0	57.1	6 683
		10:29	327.0	56.8	6 699
		10:34	331.6	68.1	6 696
		10:39	333.8	69.0	6 603

3.1 理论结论验证

在不同测量缓冲区及子测量时间段大小的条件下,采用我们的方法对 8 个数据进行流测量,在每种参数取值条件下独立运行 27 次.测量缓存取大小分别取 6 000,12 000,18 000 和 24 000,子测量时间段分别取 6s,12s 和 20s.

首先,我们计算每个流记录的分组数及字节数的相对误差,即绝对误差与实际值的比值.图 2 是某个数据 27 次测量结果的分组误差散点分布图,其中 x 轴是流的分组数(分组数大于 370),y 轴是相对误差.可以发现,随着流分组数的增加,其估计误差逐渐减小.进一步地,图形关于 y=0 水平线的对称性说明了分组数估计的无偏性.字节误差与分组数情况下的结果相同.这在实验上验证了定理 1 与定理 2 的结论.

在实际应用中,往往对具有相同属性的多个单个流记录进行汇聚从而进行对汇聚流的分析.在本文中,我们按应用层的端口号对流进行汇聚形成各种应用层汇聚流,然后求汇聚流的分组数(字节数)相对误差,其中的 P2P(peer to peer)应用包括 KAZAA,eDonkey2000 与 Gnutella,它们的端口号分别为 1 214,4 661,6 346.为了考察测量缓冲区大小对估计误差的影响,我们分析了各种汇聚流的分组数(字节数)相对误差的变化情况.如图 3 所示,当测量缓冲区变大时,汇聚流的估计误差都逐渐减小(本图中是分组相对误差,字节数相对误差情况是相同的).这与定理 3、定理 4 的结论是一致的.图 4 给出的是测量缓冲区大小固定(本图中测量缓冲区大小为 12 000)而误差随子测量时间段变化的情况:子测量时间段变大,相对估计误差也随之变大.这是因为单位时间内被抽样的分组更少.在图 5 中,上方点线中的点标识的是按定理 3 计算的各个汇聚流分组数相对误差的理论上界,呈垂直分布的散点是相应汇聚流在多次测量结果中[****]的相对误差分布.可见,这些误差低于理论上界.字节相对误差也具有相似的情况,限于篇幅,这里不再列出.

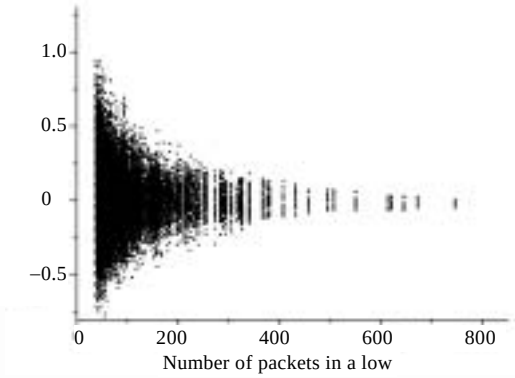

Fig.2 Scatter plot for relative error of packets total

图 2 分组数相对误差散点分布

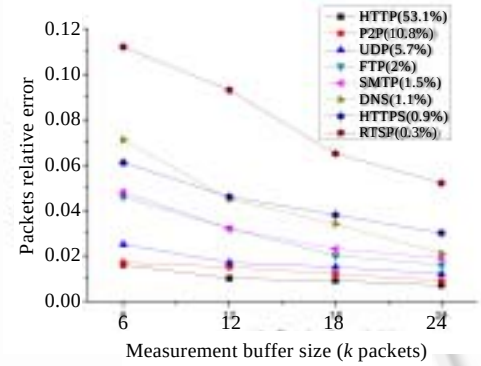

Fig.3 The relation between relative error and measurement buffer size

图 3 相对误差与测量缓冲区大小之间的关系

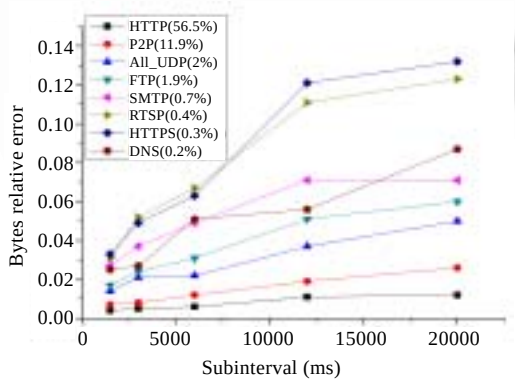

Fig.4 The relation between relative error and subinterval

图 4 相对误差与子测量时间段之间的关系

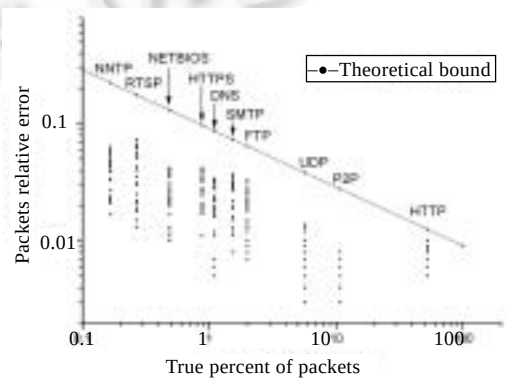

Fig.5 The relative error of packets total for aggregate flows

图 5 汇聚流分组数相对误差

**** 对 27 次测量结果进行排序后,从第 20%到第 80%之间的数据.

3.2 方法比较

本节比较我们的抽样方法 TSPSBMB 与 Cisco 随机抽样 NetFlow(RSNF)、文献[6]中 Adaptive NetFlow (ANF)的准确性.为了比较的公平性,需要保证 3 种方法所消耗的资源在相同的条件下进行.而对于同样的分组序列,抽样率直接决定了抽样方法的资源消耗量.因此,在每一个子测量时间段设置下,我们首先运行 TSPSBMB,根据最后的分组总数和抽样分组数计算一个等效抽样率;然后用这个抽样率运行 RSNF 和 ANF.根据 ANF 的抽样方法,其实质是用测量时间段结束时调整的抽样率对数据进行简单随机抽样,因此,我们用等效抽样率进行简单随机抽样来获得 ANF 的测量结果.

与上一节相同,我们使用汇聚流的估计误差来进行比较.结果表明:在等效抽样率相同的情况下,3 个抽样方法的估计误差是相当的,没有显著差别.表 2 与表 3 分别为字节数和分组数的相对误差结果比较(TSPSBMB 此时使用的测量缓冲区大小为 12 000,其中子测量时间段为 6s,12s,20s 时对应的等效采样率为 1/22,1/55,1/92).

Table 2 Comparison of relative error of bytes

表 2 字节数相对误差比较

Aggregate flows		HTTP	P2P	FTP	SMTP	RTSP	HTTPS	DNS	NETBIOS
Percent (%)		56.48	11.91	1.93	0.710	0.42	0.32	0.21	0.06
6s (1/22)	TSPSBMB	0.004	0.009	0.022	0.035	0.050	0.042	0.028	0.032
	RSNF	0.004	0.010	0.024	0.036	0.048	0.040	0.031	0.043
	ANF	0.003	0.009	0.025	0.031	0.051	0.047	0.030	0.040
12s (1/55)	TSPSBMB	0.005	0.013	0.030	0.044	0.073	0.060	0.045	0.048
	RSNF	0.007	0.013	0.033	0.052	0.071	0.068	0.041	0.052
	ANF	0.008	0.015	0.039	0.069	0.069	0.071	0.046	0.042
20s (1/92)	TSPSBMB	0.008	0.019	0.043	0.059	0.089	0.078	0.048	0.062
	RSNF	0.007	0.017	0.042	0.073	0.090	0.080	0.045	0.067
	ANF	0.008	0.020	0.037	0.082	0.085	0.079	0.049	0.071

Table 3 Comparison of relative error of packets total

表 3 分组数相对误差比较

Aggregate flows		HTTP	P2P	FTP	SMTP	DNS	HTTPS	NETBIOS	RTSP
Percent (%)		53.13	10.76	1.99	1.54	1.10	0.88	0.49	0.27
6s (1/22)	TSPSBMB	0.003	0.007	0.016	0.016	0.024	0.029	0.030	0.045
	RSNF	0.002	0.006	0.017	0.018	0.024	0.026	0.036	0.043
	ANF	0.003	0.007	0.014	0.013	0.024	0.028	0.027	0.042
12s (1/55)	TSPSBMB	0.004	0.010	0.027	0.026	0.033	0.032	0.047	0.061
	RSNF	0.004	0.009	0.025	0.028	0.032	0.033	0.049	0.068
	ANF	0.005	0.014	0.022	0.028	0.030	0.034	0.044	0.067
20s (1/92)	TSPSBMB	0.005	0.013	0.029	0.032	0.040	0.046	0.058	0.075
	RSNF	0.004	0.012	0.025	0.033	0.038	0.055	0.057	0.076
	ANF	0.006	0.012	0.031	0.040	0.039	0.054	0.058	0.073

在实际中,RSNF 由于没有自适应性和资源可控性,用户无法事先选择合适的抽样率.ANF 虽然具有自适应性,但其 CPU 的利用率较低,而且在调整抽样率时过于复杂.相对于这两种方法,我们的方法具有简单性、自适应性及资源可控性,同时也不失准确性.

4 结 语

本文针对现有 NetFlow 抽样方法的不足,提出了一种基于测量缓冲区的时间分层分组抽样方法.该方法具有实现、操控简单、参数自适应和资源消耗可控等优点.通过理论分析,我们证明了抽样估计的无偏性,并推导出估计值相对标准差的理论上界,进而揭示了影响估计误差的因素.最后,使用互联网实际数据进行了实验,验证了理论分析结论并与其他同类抽样方法进行比较.实验结果表明,我们的方法在具有简单性、自适应性及资源可控性的同时,不会失去准确性.

现有的 NetFlow 及本文使用的流超时策略都是一种固定超时策略.文献[13]提出了一种概率保证的自适应性超时算法,我们认为:使用它会进一步提高“流 cache”的利用率.另外,结合文献[12]的流抽样方法也会进一步

控制流记录输出占用的带宽资源.这些将是改进 NetFlow 的下一步工作.

致谢 我们向为本文的实验提供互联网数据的 CAIDA[16]组织及其成员 Colleen Shannon 以及对本文提出有益建议的黄晓慧和审稿专家们表示感谢.

References:

[1] Claffy KC, Braun HW, Polyzos GC. Parameterizable methodology for Internet traffic flow profiling. IEEE Journal on Selected Areas in Communications, 1995,136(8):1481−1494.

[2] Brownlee N, Mills C, Ruth G. Traffic flow measurement: Architecture. RFC 2722, 1999.

[3] Brownlee N, Plonka D. IP flow information export (ipfix). IETF working group. 2005. http://www.ietf.org/html.charters/ipfix-charter.html

[4] Cisco netflow. 2005. http://www.cisco.com/warp/public/732/Tech/netflow

[5] Random sampled netflow. 2005. http://www.cisco.com/en/US/products/sw/iosswrel/ps5207/products_feature_guide09186a00801a7618.html

[6] Estan C, Keys K, Moore D, Varghese G. Building a better netflow. ACM SIGCOMM Computer Communication Review, 2004, 34(4):245−256.

[7] Claffy KC, Polyzos GC, Braun HW. Application of sampling methodologies to network traffic characterization. ACM SIGCOMM Computer Communication Review, 1993,23(4):194−203.

[8] Drobisz J, Christensen KJ. Adaptive sampling methods to determine network traffic statistics including the Hurst parameter. In: Proc. of the IEEE Annual Conf. on Local Computer Networks. Piscataway: IEEE Press, 1998. 238−247.

[9] Cheng G, Gong J, Ding W. Network traffic sampling measurement model on packet identification. Acta Electronica Sinica, 2002, 30(12A):1986−1990 (in Chinese with English abstract).

[10] Wang JF, Yang JH, Zhou HX, Xie GG, Zhou MT. Adaptive sampling methodology in network measurements. Journal of Software, 2004,15(8):1227−1236 (in English with Chinese abstract). http://www.jos.org.cn/1000-9825/15/1227.htm

[11] Choi BY, Park J, Zhang ZL. Adaptive packet sampling for accurate and scalable flow measurement. In: Proc. of the IEEE Global Telecommunications Conf., Vol.3. New York: IEEE Press, 2004. 1448−1452.

[12] Duffield NG, Lund C, Thorup M. Flow sampling under hard resource constraints. ACM SIGMETRICS Performance Valuation Review, 2004,32(1):85−96.

[13] Wang J, Li L, Sun F, Zhou M. A probability-guaranteed adaptive timeout algorithm for high-speed network flow detection. Computer Networks, 2005,48(2):215−233.

[14] Vitter JS. Random sampling with a reservoir. ACM Trans. on Mathematical Software, 1985,11(1):37−57.

[15] Rice JA. Mathematical Statistics and Data Analysis. 2nd ed., Beijing: China Machine Press, 2003. 185−225.

[16] Cooperative Association for Internet Data Analysis. 2005. http://www.caida.org/

附中文参考文献:

[9] 程光,龚俭,丁伟.基于分组标识的网络流量抽样测量模型.电子学报,2002,30(12A):1986−1990.

[10] 王俊峰,杨建华,周虹霞,谢高岗,周明天.网络测量中自适应数据采集方法.软件学报,2004,15(8):1227−1236. http://www.jos.org.cn/1000-9825/15/1227.htm

王洪波(1975－),男,河北石家庄人,博士,讲师,主要研究领域为IP网络测量,网络管理,网络安全.

林宇(1976－),男,博士,副教授,主要研究领域为互联网服务质量管理与测量,P2P计算.

韦安明(1975－),男,博士生,主要研究领域为IP网络测量,网络安全检测.

程时端(1940－),女,教授,博士生导师,主要研究领域为互联网性能分析与服务质量控制,P2P计算,传感器网络.